Available online at www.sciencedirect.com
ScienceDirect
Applied Mathematics Letters 21 (2008) 516–520
Applied Mathematics Letters
www.elsevier.com/locate/aml

# Protected points in ordered trees

Gi-Sang Cheon[a], Louis W. Shapiro[b,*]

[a] Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Republic of Korea
[b] Department of Mathematics, Howard University, Washington, DC 20059, USA

Received 29 June 2007; accepted 20 July 2007

**Abstract**

In this note we start by computing the average number of protected points in all ordered trees with $n$ edges. This can serve as a guide in various organizational schemes where it may be desirable to have a large or small number of protected points. We will also look a few subclasses with a view to increasing or decreasing the proportion of protected points.
© 2007 Elsevier Ltd. All rights reserved.

*Keywords:* Ordered tree; Protected point; Catalan number; Fine number; Central binomial coefficient; Motzkin numbers; {0, 1, 2}-trees

## 1. Notation and overview

An *ordered tree* [2] is defined recursively. It is a tree with a root and an ordered list of subtrees at the root. For instance the subtrees could be ordered by the time of creation. The five ordered trees with three edges are shown in Fig. 1.

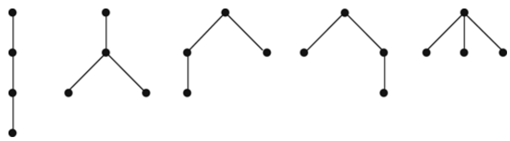

Fig. 1. Ordered trees with three edges.

A *protected point* is a vertex which is not a leaf and which is not distance 1 from a leaf. The root is not considered to be a leaf except for the tree consisting of only the root.

For instance, if leaves represent customers it may be worthwhile for many of the points in the tree to be unprotected. However if the leaves represent lobbyists or computer hackers it may be a very good thing to have many points protected. We will show that as the number of edges gets large the average proportion of protected points in all ordered trees approaches $1/6$. The tool we will use is generating functions. A reasonable variation occurs if we have an organizational tree such that the maximum number of employees directly under any one manager is at most two.

* Corresponding author.
*E-mail address:* lou.shapiro@gmail.com (L.W. Shapiro).

0893-9659/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aml.2007.07.001

If the out degree of any point is at most two then we are looking at *Motzkin* trees. The same tools can be used to show that the proportion of protected points in Motzkin trees approaches $10/27$. We look at these two cases in some detail and then mention three more cases.

Two generating functions which will use are those for the *Catalan numbers* and for the *central binomial coefficients*. They are

$$C(z) = C = \frac{1-\sqrt{1-4z}}{2z} = 1 + zC^2 = \frac{1}{1-zC} = \sum_{n\geq 0} \frac{1}{n+1}\binom{2n}{n} z^n,$$

and

$$B(z) = B = \frac{1}{\sqrt{1-4z}} = 1 + 2zCB = \frac{1}{1-2zC} = \sum_{n\geq 0} \binom{2n}{n} z^n.$$

An excellent source for background information on generating functions, Catalan numbers and Motzkin numbers is Stanley's book [6]. For information on Fine numbers see [2,3] while asymptotics are discussed in [1] and somewhat similar applications of tree structure are discussed in [4,5].

It is well known that the number of ordered trees on $n$ edges is the $n$th Catalan number, $C_n = \frac{1}{n+1}\binom{2n}{n}$. We use the terms point and vertex interchangeably. The number of vertices in any tree with $n$ edges is $n+1$ so the generating function for the number of ordered trees with a distinguished point is

$$B = \sum_{n\geq 0} (n+1) \frac{1}{n+1}\binom{2n}{n} z^n.$$

Alternatively this counts all vertices in all ordered trees. If we count leaves which are vertices of up degree 0 then we find the numbers $1, 1, 3, 10, 35, 126, \ldots$ which suggests the generating function $(B+1)/2$.

The other generating function we will need is that for the number of trees where the root is a protected point or is the empty tree. Trying small cases gives the numbers $1, 0, 1, 2, 6, 18, 57, \ldots$. In Fig. 1 where $n = 3$ we see that the two trees on the left have protected roots. Since each subtree out of the root must have one edge connecting to the root (the generating function for a single edge is $z$) and a nontrivial tree attached to this edge (with generating function $C-1$), each subtree at the root contributes $z(C-1) = z^2C^2$ and the total generating function is

$$1 + z^2C^2 + \left(z^2C^2\right)^2 + \left(z^2C^2\right)^3 + \cdots = \frac{1}{1-z^2C^2}.$$

This sequence of numbers is called the *Fine number sequence* [3] and the generating function for the sequence is denoted as

$$F(z) = F = \frac{1}{1-z^2C^2} = \frac{1}{(1-zC)(1+zC)} = \frac{C}{1+zC}. \tag{1}$$

For a good reference for this material see [6], and for asymptotic estimates the following lemma of Bender is easy to apply and very useful.

**Theorem 1.1** (*Bender's Lemma [1]*). *Suppose that $A(z) = \sum_{n\geq 0} a_n z^n$ and $B(z) = \sum_{n\geq 0} b_n z^n$ are two generating functions, and the radius of convergence of $A(z)$ is larger than that of $B(z)$. Let $C(z) = \sum_{n\geq 0} c_n z^n$ be the product $A(z)B(z)$. Suppose further that $b_{n-1}/b_n$ approaches a limit $b$ as $n \to \infty$. If $A(b) \neq 0$, then $c_n \sim A(b)\, b_n$.*

## 2. The main result

In ordered trees and in similar classes of trees the following observation holds:

$$V = LT \tag{2}$$

where $V$ is the generating function for trees with a distinguished vertex, $L$ is the generating function for trees with a distinguished leaf and $T$ is the generating function for the number of trees in the class. The way to see that this holds is to "snip" the distinguished vertex in half. This produces two trees. The first is a tree now with a distinguished leaf and a second which was the subtree growing up from the distinguished vertex.

**Theorem 2.1.** *The average portion of protected points in all ordered trees with n edges approaches* $1/6$ *as* $n \to \infty$.

**Proof.** For ordered trees we have $T = C$, $V = B$ where $C$ and $B$ are generating functions for the Catalan numbers and the central binomial coefficients, respectively, and what looks like $L = (B+1)/2$. Assuming this from (2) we would have

$$B = \frac{B+1}{2} \cdot C$$

or equivalently $2B = C(B+1)$. To prove this we write the right hand side as

$$\frac{B+1}{2} \cdot C = \frac{1}{2} \cdot \left( \frac{1}{\sqrt{1-4z}} + 1 \right) \cdot \left( \frac{1-\sqrt{1-4z}}{2z} \right)$$

and simplify. This does provide a proof that $L = (B+1)/2$.

If we want the distinguished point to be protected we want the tree on top to be nontrivial and to have its root protected. Thus the appropriate generating function is $F - 1$ where $F$ is the generating function for the Fine numbers given by (1). Hence we have

$$L(F-1) = \frac{B}{C}\left(\frac{C}{1+zC} - 1\right) = \frac{B}{1+zC} - \frac{B}{C}.$$

After expressing $C$ and $B$ in terms of $\sqrt{1-4z}$ in this equation we obtain

$$L(F-1) = \frac{1}{4} \cdot \frac{1}{1+\frac{z}{2}} - \frac{1}{2} + \frac{1}{\sqrt{1-4z}} \cdot \frac{1-z}{2z+4}. \tag{3}$$

The first few terms of $L(F-1)$ are

$$z^2 + 3z^3 + 11z^4 + 40z^5 + 148z^6 + 553z^7 + 2083z^8 + 7896z^9 + O(z^{10}).$$

Now, we are ready to find asymptotic values. Asymptotically the first two terms of the right hand side in (3) are irrelevant and using Bender's lemma with $A(z) = \frac{1-z}{2z+4}$, $B(z) = \frac{1}{\sqrt{1-4z}} = \sum_{n\geq 0} \binom{2n}{n} z^n$ and $b = \frac{1}{4}$, we see that

$$[z^n]\frac{1}{\sqrt{1-4z}} \cdot \frac{1-z}{2z+4} \to \binom{2n}{n} \frac{1-1/4}{2(1/4)+4} = \frac{1}{6}\binom{2n}{n},$$

where $[z^n]$ is the coefficient operator. Since the total number of points is $\binom{2n}{n}$ we have that the average number of protected points approaches $1/6$. ■

For numerical reassurance we note that

$$\frac{[z^{50}]L(F-1)}{[z^{50}]B} = \frac{16739992778065482809017276636}{100891344545564193334812497256} \doteq 0.16592.$$

Does a system where each staff member can hire at most two underlings afford a higher percentage of protected points? To determine this we look at $\{0, 1, 2\}$-trees where the out degree of every vertex is 0,1, or 2. The numbers of these trees are counted by the Motzkin numbers $M_n$ with the generating function

$$M(z) = M = \frac{1-z-\sqrt{1-2z-3z^2}}{2z^2}. \tag{4}$$

It is also known that the number of all vertices in $\{0, 1, 2\}$-trees, i.e. the number of $\{0, 1, 2\}$-trees with a distinguished vertex, has the generating function $V$ given by

$$V = \sum_{n\geq 0}(n+1)M_n z^n = \frac{\mathrm{d}}{\mathrm{d}z}(zM)$$

since any tree with $n$ edges has $n+1$ vertices. But then, after some manipulation,

$$V = \frac{\mathrm{d}}{\mathrm{d}z}(zM) = \frac{1-z-\sqrt{1-2z-3z^2}}{2z^2} \cdot \frac{1}{\sqrt{1-2z-3z^2}}.$$

Since $V = LT$ we see that the number of $\{0, 1, 2\}$-trees with a distinguished leaf has the generating function

$$L = \frac{V}{T} = \frac{1}{\sqrt{1-2z-3z^2}}. \tag{5}$$

We note that $L$ has a singularity at $z = 1/3$ so the radius of convergence about $z = 0$ is also $1/3$ and the ratio test then tells us that

$$\lim_{n\to\infty} \frac{l_{n+1}}{l_n} = 3, \quad \text{where } l_n = [z^n]L. \tag{6}$$

To get our asymptotic result we now express everything in terms of $l_n$. Since $V$ may be rewritten as

$$V = \frac{1-z}{2z^2\sqrt{1-2z-3z^2}} - \frac{1}{2z^2},$$

asymptotically we have

$$\begin{aligned}\left[z^n\right] V &= \left[z^n\right] \frac{1-z}{2z^2\sqrt{1-2z-3z^2}} = \frac{1}{2}\cdot(l_{n+2}-l_{n+1}) \\ &\sim \frac{1}{2}\cdot(9l_n - 3l_n) = 3l_n \quad \text{as } n\to\infty.\end{aligned} \tag{7}$$

This is a result of some independent interest since it tells us that for $\{0, 1, 2\}$-trees as $n$ gets large about $1/3$ of the vertices are leaves.

**Theorem 2.2.** *The average portion of protected points in* $\{0, 1, 2\}$*-trees with n edges approaches* $10/27$ *as* $n \to \infty$.

**Proof.** By a similar argument, the number of $\{0, 1, 2\}$-trees where the root is a protected point or the empty tree has the generating function

$$K := 1 + z(M-1) + z^2(M-1)^2,$$

where $M$ is the generating function for the Motzkin numbers given by (4). Thus, the generating function for the number of protected points on $\{0, 1, 2\}$-trees is given by

$$\begin{aligned}L(K-1) &= L(z(M-1)+z^2(M-1)^2) \\ &= \frac{(1-2z^2)M}{\sqrt{1-2z-3z^2}} + \frac{z^2-z-1}{\sqrt{1-2z-3z^2}} \\ &= \frac{2z^4-4z^2-z+1}{2z^2\sqrt{1-2z-3z^2}} + 1 - \frac{1}{2z^2}.\end{aligned} \tag{8}$$

The first few terms of $L(K-1)$ are

$$z^2 + 3z^3 + 10z^4 + 31z^5 + 94z^6 + 281z^7 + 834z^8 + 2465z^9 + O(z^{10}).$$

Asymptotically the last two terms of the right hand side in (8) are irrelevant and

$$\begin{aligned}[z^n]L(K-1) &= [z^n]\frac{2z^4-4z^2-z+1}{2z^2\sqrt{1-2z-3z^2}} \\ &= \frac{1}{2}(2l_{n-2} - 4l_n - l_{n+1} + l_{n+2}) \\ &\sim \frac{1}{2}\left(2\cdot\frac{1}{9}l_n - 4l_n - 3l_n + 9l_n\right) = \frac{10}{9}l_n \quad \text{as } n\to\infty.\end{aligned} \tag{9}$$

From (7) and (9), we can now estimate the average number of protected points in {0, 1, 2}-trees as $n \to \infty$:

$$\frac{[z^n]L(K-1)}{[z^n]V} \sim \frac{\frac{10}{9} l_n}{3 l_n} = \frac{10}{27} = 0.37037. \quad \blacksquare$$

We note that

$$\frac{[z^{100}]L(K-1)}{[z^{100}]V} = \frac{2703138330664648759261590946527881993933801848 2}{7447897271050759943050224248101637348052367056 9} \doteq 0.36294.$$

Here are three more structures but we now omit details since the method is the same.

Another protocol might be that everyone has two employees if they have any at all. This situation is modeled by complete binary trees and we have in that case that $T = C$, $L = B$ and $V = \sum_{n \geq 0} \binom{2n+1}{n} z^n$. The ratio of protected points to all points is

$$\frac{n^2 - 3n + 3}{4n^2 + 3n} \to \frac{1}{4} \quad \text{as } n \to \infty. \tag{10}$$

Yet another protocol is that everyone can hire a junior or a senior employee or both but the two positions are different. The model is incomplete binary trees and we have $T = C^2$, $L = B$ and $V = \sum_{n \geq 0} \binom{2n+2}{n} z^n$. This time the ratio of protected points to all points is

$$\frac{\binom{2n}{n-2} - 2\binom{2n-2}{n-2} + \binom{2n-4}{n-2}}{\binom{2n+2}{n}} \frac{n^2 - 3n + 3}{4n^2 + 3n} \to \frac{9}{64} \cdot \frac{1}{4} = \frac{9}{256} \quad \text{as } n \to \infty. \tag{11}$$

For our last example we consider complete ternary trees where the out degree of every vertex is 0 or 3. Using similar methods we find that the ratio of protected points to all points is

$$\frac{n^3 - 6n^2 + 11n - 6}{81n^3 - 54n^2 - 9n + 5} \to \frac{1}{81} \quad \text{as } n \to \infty. \tag{12}$$

This is a very small number of protected points.

Many variations are possible using the same tools and the method is more important than any particular case. The method may be thought of as taking four steps:

1. Find the number of protected points at the root for the class of ordered trees being considered.
2. Use the $V = LT$ equation to find the generating function $L$.
3. Use $L$ to transport the generating function at the root to an arbitrary vertex.
4. Find the asymptotic value, often by using Bender's lemma or at least the radius of convergence of the relevant generating functions.

## Acknowledgments

We would like to thank the referee for thoughtful and helpful suggestions. The second author was supported in part by NSF Grant HRD-040 1697.

## References

[1] E. Bender, Asymptotic methods in enumeration, SIAM Rev. 16 (1974) 485–515.
[2] E. Deutsch, L.W. Shapiro, A bijection between ordered trees and 2-Motzkin paths and its many consequences, Discrete Math. 256 (2002) 655–670.
[3] E. Deutsch, L.W. Shapiro, A survey of the Fine numbers, Discrete Math. 241 (2001) 241–265.
[4] J.D. Farley, Breaking Al-Qaeda cells: A mathematical analysis of counterterrorism operations (A guide for risk assessment and decision making, Stud. Conflict & Terrorism 26 (2003) 399–411.
[5] R.J. Wilson, J.J. Watkins, Graphs, an Introductory Approach, Wiley, 1990, pp. 47–59.
[6] R. Stanley, Enumerative Combinatorics, vol. 2, Cambridge University Press, 1999 (Chapter 6, Exercises 19 and 38).